November 30, 2020

2021 Alvarez-Hopper Selection Committee
Lawrence Berkeley National Laboratory
Berkeley, CA, 94720

Dear Alvarez-Hopper Selection Committee,

I am writing to enthusiastically apply for the 2021 Alvarez postdoctoral fellowship. Through my work as an engineer at Google, my PhD in statistics at Berkeley, and my current postdoctoral position at the Massachusetts Institute of Technology, my academic ambition has consistently been to bring conceptually simple, theoretically sound, and computationally tractable tools to bear on the large scale data science problems of the twenty-first century. The US Department of Energy (DOE), and the Berkeley Lab in particular, are currently aggressively pursuing new frontiers in data science, as expressed, for example, by the Advanced Scientific Computing Advisory Committee Subcommittee on AI/ML report of September 2020. The Lab's expressed need for scalable and reliable uncertainty quantification and model interrogation are a perfect fit for my research, and I am confident that close collaboration between Berkeley Lab scientists and myself would produce both valuable science and methodological advances in data science.

My work on large-scale data problems, both at Google and on the Celeste project (producing astronomical catalogs on the NERSC supercomputing cluster), have acquainted me with the difficulty of applying traditional statistical ideas at large scale. My research has been devoted to easing these difficulties. My PhD began with an effort to overcome the fact that classical Bayesian Markov Chain Monte Carlo (MCMC) techniques would not scale to the problems I encountered at Google. This research question led me to develop the linear response correction for variational Bayes approximations, a perturbative correction which provides reliable, approximate Bayesian inference on problems far too massive for MCMC (like the construction of astronomical catalogs). During the second half of my PhD, I found that similar perturbative ideas provide scalable solutions to a wide array of practical machine learning problems involving uncertainty quantification or model checking, including cross validation, prior sensitivity, adversarial data sensitivity, and sensitivity to prior or model specification.

Good ideas are made better by being put into practice, and I believe that the my research will be best developed and advanced in the context of massive, practically relevant scientific applications. In addition to revisiting the production of astronomical catalogs in light of my variational Bayes work, both the High Energy Physics and Nuclear Physics groups of Berkeley Lab have expressed explicit need for uncertainty quantification as part of the "AI for Science Initiative" in the above-mentioned report. As the report states, "While AI algorithms have proliferated, general techniques for understanding their accuracy and stability in complex environments have not." I believe that collaboration between statisticians such as myself and scientists is the best way to fill this need, and it is for this reason that I am applying to become an Alvarez fellow.

Sincerely,

Ryan Giordano